Smart Designs for Smart Technologies: Research Challenges and Emerging Solutions for Scientist-Practitioners within e-Mental Health

Abstract

Consumers are increasingly likely to access various forms of e-mental health and there is considerable danger that they may be exposed to untested interventions. Traditional research designs, such as the Randomised Controlled Trial (RCT) are limited in their capacity to match the pace of development and evolving nature of e-mental health. There are a number of unique challenges associated with research into the development and use of technologically based interventions. This paper provides a discussion of these challenges, and examines emerging strategies that may enable clinicians to have more confidence in integrating e-mental health in their practice. We argue that greater use of small sample size designs, greater collaboration and research in applied settings, as well as more focused empirical investigation during program development stages is needed. We use a research example of a Smartphone application (app) aimed at the treatment of anxiety disorders to illustrate the procedure, value and clinical applications of each of the emerging research designs.

Keywords: e-mental health, mHealth, single case research, Sequential Multiple Assignment Randomized Trial, Multiphase Optimisation Strategy

Smart Designs for Smart Technologies: Methodological Challenges and Emerging Solutions for Scientist-Practitioners within e-Mental Health

As we noted in an earlier paper published in Professional Psychology: Research and Practice (Clough & Casey, 2015a), the exponential growth in e-mental health services is well documented (Clough & Casey, 2011a, 2011c, 2015a; Griffiths, Farrer, & Christensen, 2007; Tomlinson, Rotheram-Borus, Swartz, & Tsai, 2013) and offers considerable promise in overcoming many barriers to accessing mental health care, such as cost, location, availability of services, and stigma (Casey, Joy, & Clough, 2013; Casey, Wright, & Clough, 2014). Yet there is a considerable gap between uptake of these programs and evidence to support their effectiveness (Donker et al., 2013). In 2013, over 10,000 mental health related Smartphone Applications (Apps) were available for consumer download (Ben-Zeev et al., 2013). However, a comprehensive search of the literature in the same year (Donker et al., 2013) identified only five Apps that had been tested using either a control group or pre-to-post design. As consumers increasingly engage with these technologies, clinicians need to be able to recommend ways that e-mental health can be incorporated in treatment and ongoing maintenance of mental health.

We have discussed some of the unique challenges associated with research in this field and how these challenges have slowed uptake of technologically based approaches and adjuncts into clinical practice. (Clough & Casey, 2015a). The purpose of the current paper is to expand on this discussion as well as examine emerging strategies and designs for overcoming these challenges in clinical practice.

1. Limitations and Difficulties in e-Mental Health Research

There are a number of challenges in combining technologically based research and clinical practice. As a subsection of e-mental health, mHealth is a rapidly expanding area of potential engagement with consumers, and is defined as any psychological or mental health intervention that is delivered or supported by the use of mobile technology (Clough & Casey, 2015a). Despite the rapid consumer uptake of mHealth technologies, research in this field is often underpowered or lacking methodological rigour (Olff, 2015).

The Randomised Controlled Trial (RCT) is considered to be the "gold standard" of research designs to examine efficacy and infer causality. In seeking guidance for evidence-based interventions with their clients, clinicians typically place most confidence in RCT results. However, there is a striking lack of congruence between the RCT and the research pace required for technological interventions to remain relevant (Clough & Casey, 2015a). Ioannidis (1998) identified the median time from grant application to publication for RCTs was 7 years, with 5.5 years between initial enrolment and publication. During this time frame, it is likely that technology may become superseded or obsolete, such as when considering the progression from personal digital assistant to Smartphone (Kumar et al., 2013). The first Smartphone designed for widespread adoption was released in 2002 (Blackberry), yet in the 13 years since this time the technological advances have been substantial. An even more serious problem concerns the length of time associated with the progression from research to clinical practice. It has been estimated that from project concept to community implementation of techniques takes on average 17 years (Riley, Glasgow, Etheredge, & Abernethy, 2013). Although the therapeutic approaches utilised in e-mental health interventions often have an established evidence base, use of the traditional RCT approach to examine the efficacy of the new

technology as a delivery modality may still result in a considerable time lag between completion of evaluation and release to consumers.

RCTs not only pose problems in terms of the associated time lag, but are also costly and create additional challenges in terms of randomisation of clients and the level of treatment adherence required (Kumar et al., 2013). The inclusion/ exclusion criteria employed in RCTs can at times lead to high levels of internal validity, but poor external validity when programs are eventually deployed in community settings. For some research designs, the client characteristics, level of comorbidity, diagnostic severity, and treatment adherence may be considerably different outside of the research protocol.

Problems can also arise during an RCT due to the ongoing revisions often required for a technological intervention. For example, the automatic software updates often included with Smartphone devices can create difficulties in a previously programmed Application. Although not all updates substantively affect the app, this software update is beyond the research team's control and may require parts of the App to be reprogrammed. A traditional RCT design requires that the intervention remains unmodified for the trial period. However it is clear that without modification the intervention may no longer being delivered consistently across time, or continuing to test the true efficacy of the intervention. Mohr (2009) describes this state as "perpetual beta" whereby the system is continuously changing. The challenge is then how to evaluate this system effectively.

Riley et al. (2013) argue that research needs to be more rapid, responsive and relevant. They recommend replacing traditional pilot testing with iterative N=1 designs, collaboration with community and industry stakeholders to enable speedier

recruitment of clients and subsequent adoption of practices, and optimisation strategies such as sequential multiple assignment.

Current research demonstrates that consumers are not waiting for scientific validation before adopting these interventions, particularly with regards to mHealth technologies (Kumar et al., 2013). Creative research designs are required in order for clinicians to make the optimal use of technological advances (Glasgow, Magid, Beck, Ritzwoller, & Estabrooks, 2005). The goal of these strategies is to ensure that both clients and clinicians may be confident that empirically supported interventions are being offered.

2. Emerging Solutions

In order to outline the emerging designs and strategies in this area we use one of our ongoing research projects to illustrate with each design how it might be applied to answer various clinical questions.

2.1. PsychAssist – An mHealth Example

PsychAssist is a Smartphone application we have designed to assist in the treatment of anxiety in adults (Clough & Casey 2015b). The application contains various homework exercises, handouts, and activities based on a primarily cognitive-behavioural approach to treatment. Content included in the App focuses on psychoeducation, motivation, monitoring, cognitive flexibility, mindfulness, and exposure. As described in Clough & Casey (2015e) initial feedback from users was highly positive, with many clients expressing a desire to use Smartphones as a part of the their psychological treatment. However, this initial evaluation of PsychAssist

identified the need for further reprogramming and refinement, requiring consideration of the various ways in which the new version of the App may be tested.

2.2. Single Case Research

Although Single Case Research (SCR) or *N =1* designs have been recognised for many years, the increasing emphasis placed on large-scale group designs has substantially diminished the use of SCR in the field (Nock, Michel, & Photos, 2007). Given that SCR can not only provide an estimate of treatment effects, but also infer causality (particularly if an iterative multiple baseline approach is employed; Kazdin & Kopel, 1975; Nock et al., 2007), these designs should be given greater consideration within the e-mental health field. SCR can be rapidly implemented, and provide results more consistent with the person centred approach to treatment (Riley et al., 2013), which may be of particular advantage to clinicians who find standard RCTs to not be representative of their typical client group. Furthermore, Riley et al. (2013) argue that with the use of Bayesian techniques to combine data from a series of SCRs there can be sufficient evidence of generalisability, reducing the need for a larger trial.

In SCR, each client acts as their own control, rather than being compared to another individual or group (Lillie et al., 2011). As such, the effects of interventions may be determined by analysing repeated measurements over time (Lillie et al., 2011). The intervention is typically deployed in phases, which may include a baseline or control phase (A) and an intervention phase (B) delivered in varying sequences (e.g., AB, ABA, ABAB, etc.; Barlow & Hersen, 1973). Many texts already exist describing SCR designs and considerations (e.g., Almirall, Nahum-Shani, Sherwood,

& Murphy, 2014; Lillie et al., 2011). However, a more recent development in this area concerns the statistical procedures for SCR.

Analysis within SCR typically takes one of two approaches: trend analysis or dominance analysis. Trend refers to the tendency for scores to move in a similar direction over time, whilst dominance refers to the superiority of one phase over another (Parker, Vannest, & Davis, 2011). By using trend analysis one can determine whether the client's scores improve *within* the intervention phase (B) or remain stable *within* the control phase (A). However, this approach fails to determine whether the difference *between* the phases is of significance. Conversely, dominance, or non-overlap analysis, examines between phase differences but is insensitive to within phase trend. That is, to what extent does the data in phase B fail to overlap (on the Y/ dependent axis) with the data in phase A? For example, the analysis can be used to determine the extent of (non)overlap in a client's anxiety scores between baseline (phase A) and intervention (phase B) phases of treatment.

In many instances, it is of interest to address questions related to both trend and dominance. It may be important to determine firstly if the intervention phase is overall superior to the control phase, and secondly if there is trend within the intervention phase such that the client is demonstrating improvement over time. Until recently, trend and dominance were addressed by separate statistical analyses, most of which require that data fit specific assumptions such as normality, linearity, and measured on at least an interval scale. These traditional analysis techniques may also be sensitive to phase A trends (instability of control phase) and autocorrelation (serial dependency associated with the fact that all responses are produced by the same person; Bengali & Ottenbacher, 1998). A statistical technique recently developed by Parker et al. (2011) overcomes many of these obstacles.

2.2.1. Tau-U. Parker et al. (2011) have proposed the statistic Tau-U as a technique to test both trend and dominance in SCR. Tau-U is based on Mann Whitney U and Kendall's Rank Correlation (Parker, Vannest, & Davis, 2011). It is a distribution free measure that does not require interval scale measurement, linearity, or constant variance.

Parker et al. (2011) field tested the statistical approach on a sample of 382 published articles reporting on SCR. They found that Tau-U is robust to the influence of outlier scores (of particular importance in client based research in which weekly fluctuations in scores may be common), performs well with autocorrelated data, and has greater statistical power than other dominance approaches. It also allows for the control of phase A trends. However, the key strength of this analytical approach is in its ability to allow for trend analysis, dominance analysis, or a combination of both through the one statistical procedure.

2.2.2. An Applied Example. To illustrate the potential use of Tau-U, we apply the approach to the example of PsychAssist. Following a period of reprogramming, we are interested in conducting further pilot testing of the App, and to do so in a way that is both expedient and empirical. An iterative multiple baseline design is chosen, whereby four clients suffering anxiety disorders engage in a control baseline period (A) and intervention period (B) using the PsychAssist App (AB design). Prior to commencing data collection, standard informed consent procedures are employed, with clients agreeing to the de-identified analysis and presentation of their individual data as well as to be randomly allocated to baseline conditions. The length of baselines is varied so that the intervention period is staggered, allowing for greater inference of causality. Clients are randomly allocated to one of 3, 4, 5, or 6-week baseline periods as illustrated in Table 1, following which they complete the 9-

week therapy program. The Depression Anxiety and Stress Scales-21 item version (DASS21; Lovibond & Lovibond, 1995) is administered weekly, with a focus on change in anxiety as a result of the intervention.

Table 1

Design for Multiple Baseline SCR Pilot

Client	Measurement Point														
1	A1	A2	A3	B1	B2	B3	B4	B5	B6	B7	B8	B9			
2	A1	A2	A3	A4	B1	B2	B3	B4	B5	B6	B7	B8	B9		
3	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	B6	B7	B8	B9	
4	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	B7	B8	B9
	Phase A (baseline control)						Phase B (intervention)								

A standard SCR approach is utilised, with each case analysed separately. A set of simulated data for Client 4 is displayed in figure 1, using the dependent variable of score on the anxiety subscale of the DASS-21. The free online Tau-U calculator (http://www.singlecaseresearch.org/calculators/tau-u) was used to conduct trend and dominance analysis. Data was imputed and checked for Phase A trend. As Figure 1 displays, the client's anxiety started to decrease during the baseline period, perhaps due the effects of time or treatment expectations. This baseline trend was significant (Tau-U = -0.80, p = .024).

To examine the effect of the treatment program, we are interested in data improvement over time considering both non-overlap (dominance of one phase over the other) and phase B trend (the tendency for scores to improve over the treatment period). In this weighted contrast, the effect of Phase A trend will also need to be controlled. The online Tau-U calculator has been purpose built to identify and enable control of unwanted baseline effects. After controlling for the unwanted baseline

effects, the weighted effect of non-overlap between the two phases and phase B trend was significant (Tau-U = -0.70, $p < .001$), such that the treatment program resulted in statistically significant reductions in anxiety for Client 4. That is, a 70% overall improvement between phases and during treatment (after controlling for baseline trend effects) was observed.

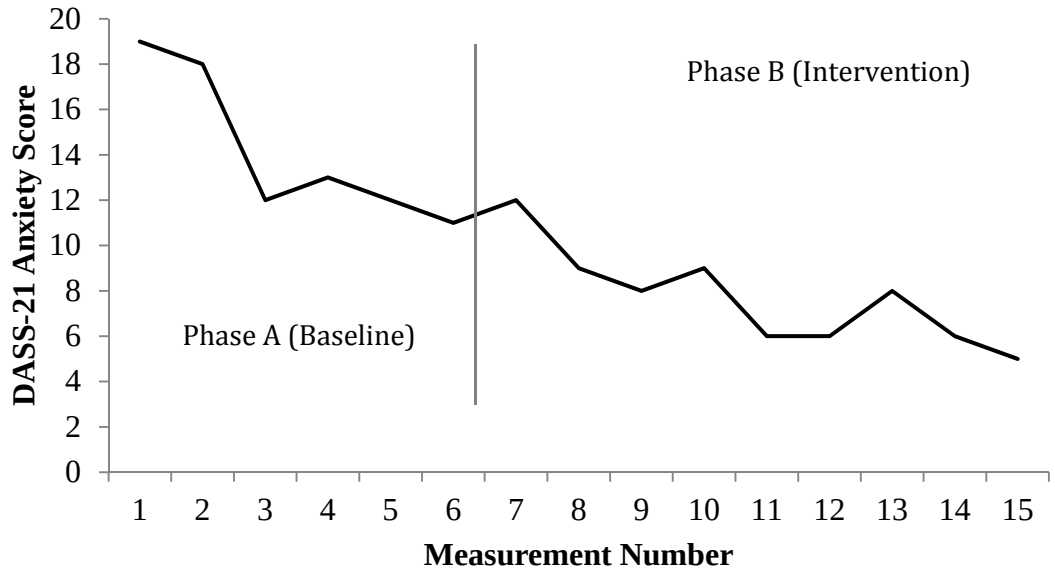


Figure 1. Simulated data for Client 4 on the DASS-21 anxiety subscale.

SCR analysis should not only include statistical analysis but visual analysis, and depending on the measure, an analysis of reliable and clinically significant change. To examine the overall effects of the intervention, case contrasts can be combined in the online calculator to produce an overall weighted effect size (Parker, Vannest, Davis, et al., 2011). A preliminary conclusion regarding the efficacy of treatment using PsychAssist may then be drawn based on the individual and overall effects. These types of analyses previously involved considerable time and specialised knowledge to learn multiple analyses and decision criteria but with advances such as

with Tau-U, are much easier to learn and conduct enabling clinicians to focus on interpretation of the data.

2.2.3. Population Estimates. One of the main criticisms of SCR designs has been that they lack in external validity and do not assess the population level effects of treatments (Lillie et al., 2011). However, developments in the statistical techniques used in this area now mean that SCRs have much greater powers of generalisability. A number of techniques allow effect sizes to be examined across studies by means of meta-analysis for SCR (e.g. Manolov, Guilera, & Sierra, 2014; Shadish, 2014; Zucker, Ruthazer, & Schmid, 2010). Combining the effect sizes across SCR studies can give population treatment effect estimates, considerably enhancing the research capacity associated with SCR designs (Zucker et al., 2010). Using PsychAssist as an example, collaboration with community partners may enable a number of SCRs may be conducted across settings with different client groups and therapists. Effect sizes from these different SCR studies may then be combined using meta-analysis to estimate the population level effect of the treatment (Busse, Kratochwill, & Elliott, 1996; Faith, Allison, & Gorman, 1996).

2.2.4. Summary. SCR has the capacity to overcome many of the challenges associated with traditional research designs, including the time associated with the conduct and dissemination of trials. SCR allows for the timely and empirical examination of intervention effects, allowing for more rapid dissemination of results and promoting rigour from the pilot testing stage onwards. Furthermore, SCR may also facilitate greater research to be conducted in applied settings and for clinicians to be able to directly monitor the effects of technologies on each individual client. These designs also typically require fewer resources than standard RCTs and may thereby require less funding. The collaborative use of these designs with industry and

healthcare providers may also lead to more efficient use of resources and increased competitiveness in accessing resources.

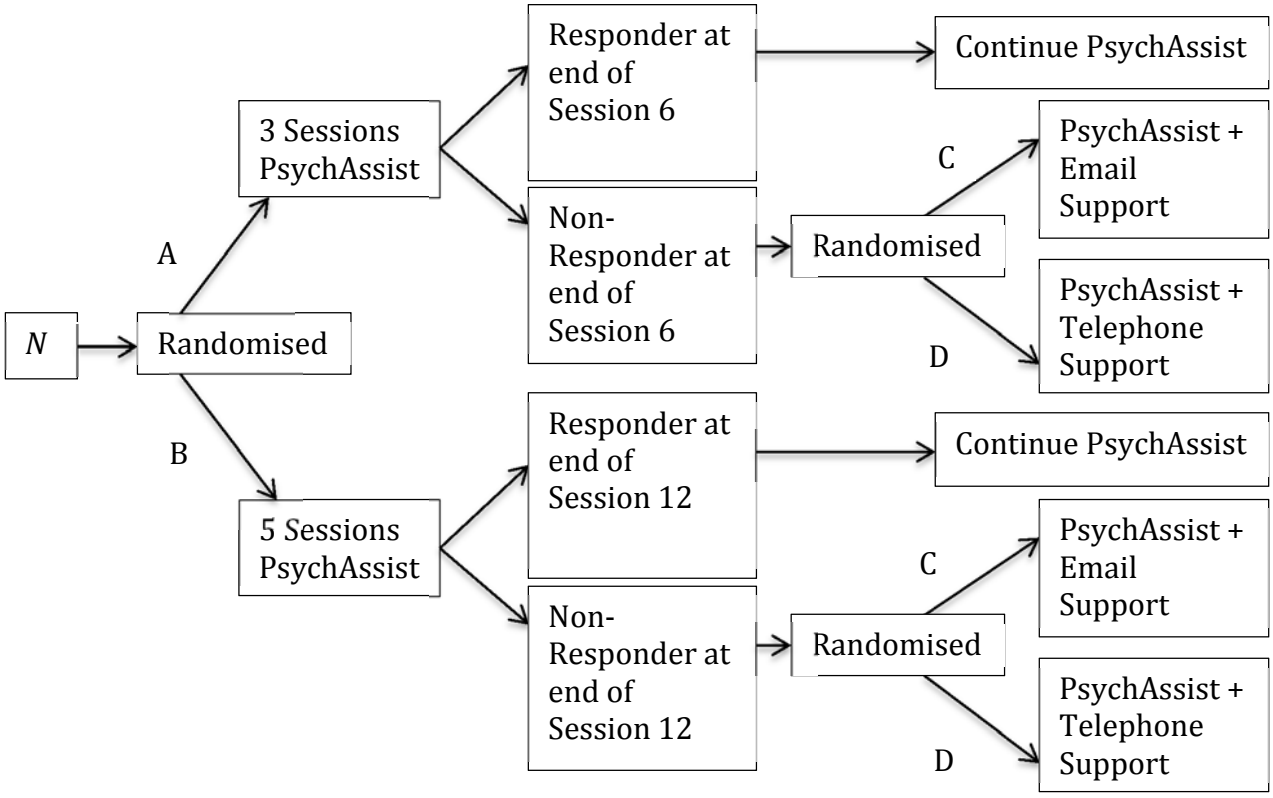
2.3. Sequential Multiple Assignment Randomized Trial (SMART)

SMART is a recently developed research design, which has been coined as an "adaptive intervention" (Almirall et al., 2014). SMART designs allow for the examination of individualised treatment effects (Shortreed et al., 2011). As noted previously, one difficulty associated with RCTs is that the intervention is examined as a whole, and requires that each client in the intervention be delivered the same treatment and follows the same process through the research protocol. However, when these interventions are applied in real world settings, treatment is typically tailored and adapted over time to meet the specific needs and treatment response of the individual (Almirall et al., 2014). The rules that clinicians use to judge whether a treatment needs to be adapted (and for which individuals) are often not based on research evidence but experience or clinical intuition.

SMART, as an adaptive intervention, is based on a sequence of decision rules that specify "whether, how, when (timing), and based on which measures, to alter the dosage (duration, frequency or amount), type, or delivery of treatment(s) at decision stages in the course of care" (Almirall et al., 2014, p. 262). The goal of the SMART approach is therefore to understand the best sequencing of intervention components (Collins, Murphy, & Strecher, 2007). Within e-mental health, this approach can answer a number of important treatment questions, such the optimal length of an intervention, the best approach to take with treatment non-responders, and the level of support required for individuals. SMART designs can ensure that the final e-mental health program that is delivered, is in fact the most powerful version of the intervention (Riley, Serrano, Nilsen, & Atienza, 2015).

Within a SMART design, four elements are identified: decision stages (such as when treatment will be assessed and tailored), treatment options (which treatments or dosages are being assessed and at which stages), tailoring variables (information about the individual that is used in making treatment decisions), and decision rules (links the tailoring variable to the treatment option) (Almirall et al., 2014). Clients are then randomly allocated within the SMART design to test the efficacy and effectiveness of the decision rules.

2.3.1. An Applied Example. A SMART design may be applied to further test the PsychAssist App. For the next step in investigating this App, suppose that we have questions regarding the optimal amount of support required (PsychAssist alone, or Psych Assist with email or telephone support) and when this support should be received (from the third or fifth session) in the nine-session program. However, when considering the amount of support required for the intervention, it is likely that this



will be dependent on the individual. We can design a SMART to examine these questions, as displayed in Figure 2.

Figure 2. Example SMART design for PsychAssist

In this example the total number of clients (*N*) is randomly allocated to receive either three (pathway A) or five (pathway B) sessions of therapy with PsychAssist. At the end of the initial treatment period clients are assessed for response to treatment (the tailoring variable). In this example, response to treatment could be determined by whether the client has achieved reliable and clinically significant change over the treatment period. Responders continue with the assigned treatment path (PsychAssist for the remainder of the nine sessions), while non-responders receive additional support. The non-responders are then randomly allocated again, to receive additional support either by email (pathway C) or telephone (pathway D) for the remainder of the program.

The design would enable us to answer the initial research questions regarding the optimal level of support for the program and when that support should be delivered. Firstly, by assessing the main effect of the first stage of treatment (all clients in pathway A compared to all clients in pathway B) we can determine when is the best time for the adaptive intervention to begin. That is, did clients do better with an early or late adaptive intervention, or was there no difference in relation to when the support was deployed?

At the second stage of treatment, a further analysis of main effects can determine the most effective level of support for the program. That is, regardless of the timing of the adaptive intervention, did the email or telephone support produce better results for PsychAssist? This analysis would involve a comparison between

pathway C (combined) and pathway D (combined). As these analyses compare across conditions, the power required is only that of a two group comparison (Collins et al., 2007). Furthermore, by assessing the interactive effect between the conditions (time of initial assessment x type of support) we can determine which of the four adaptive interventions resulted in the greatest reduction in symptomology.

2.3.2. Summary. SMART designs provide a framework to empirically determine the most powerful version of an e-mental health intervention, as well as the best way to tailor the intervention to individuals (Almirall et al., 2014; Collins et al., 2007). SMARTs can be designed and analysed in such a way as to answer more complex clinical questions, such as which intervention should be deployed based on a person's level of readiness to change, and how and when the treatment should be tailored based on their progression through the stages of change (Rivera, Pew, & Collins, 2007). However, it should be noted that the SMART design does not compare the intervention to a control or comparative treatment condition. That is, it will assist in determining the optimal version of the intervention, which would then need to be tested against a comparative treatment or control condition (Collins et al., 2007).

SMART designs may have benefits for clinicians and policy makers. SMART designs may be used to determine the optimal level of treatment (e.g., e-mental health with no therapist assistance/ minimal therapist assistance/ moderate therapist assistance) based on individual characteristics at baseline (e.g., severity, previous contact with services, readiness to change) and throughout treatment (e.g., response to treatment). This would provide clinicians with empirically supported guidance in making treatment decisions, rather than relying on experience or intuition.

Furthermore, these designs may be used to guide policy regarding stepped care approaches to treatment.

2.4. Multiphase Optimisation Strategy (MOST)

MOST is a strategy for determining the most effective version of an e-mental health intervention (Collins et al., 2011; Collins et al., 2007). It is a system aimed at creating optimal versions of multicomponent interventions. That is, MOST is an approach to conducting research, rather than a specific design. The design of MOST experiments will vary depending on the research questions under consideration (Collins et al., 2011). However, it can be thought of as a method for testing the relative effects of components within a treatment program.

Research conducted with the use of MOST consists of four steps. The first step involves the identification and selection of intervention components (Collins et al., 2011). Selection may be based on clinician experience, theory, or evidence-based interventions from other delivery formats (Tomlinson et al., 2013). In the second step, components are examined using randomised experimentation procedures. That is, clients are randomly allocated to conditions examining different components or dosages of the intervention. This step is used to determine which components should be included in the intervention, and at what levels or doses. The third step of MOST involves a period of fine tuning and refinement to produce the final intervention program. The final step consists of examination of efficacy and effectiveness by means of standard trial procedures (Collins et al., 2011).

In assessing the relative effects of intervention components (step 2), Collins et al (2011) suggest the use of individual experimental procedures (comparing one group to another), full factorial procedures (allows for the examination of interactions), or

fractional factorial procedures. Collins et al. argue that the use of MOST will result in the most optimal version of treatment interventions going forward to efficacy and effectiveness testing.

2.4.1. An Applied Example. Returning to the example of PsychAssist, a MOST design may be used to determine which treatment components should be included in the mHealth App and at what dosages. For example, PsychAssist currently contains five broad components; motivation enhancement, mindfulness, cognitive restructuring, emotion exposure, and relapse prevention. Based on previous research, mindfulness, cognitive restructuring and exposure are identified as core modules. As such our focus is on the motivation enhancement and relapse prevention modules (step 1). Suppose that we have three main treatment questions: 1) should the motivation enhancement module be included in the intervention, 2) should the relapse prevention module be included in the intervention, and 3) what dose of relapse prevention should be administered?

These questions can be answered through the use of MOST. Clients are randomly allocated to differing levels of motivation enhancement (present or absent), relapse prevention (present or absent), and dose of relapse prevention (one, two or three sessions) within the intervention (step 2). This gives rise to a 2 X 2 X 3 design. The effects of the modules can be tested using simple t tests to answer each question, or a factorial design (full or fractional) to also examine interactive effects. Results of the testing are used to create a refined version of the App that delivers the intervention at optimal levels (step 3). The refined version of the App is then tested for efficacy and effectiveness (step 4), which Collins (2011) recommends achieving by means of an RCT.

2.4.2. Summary. MOST is a method for empirically examining and developing optimal multicomponent e-mental health interventions (Collins et al., 2007). MOST can be used to ensure that only those modules contributing to treatment effectiveness are included in interventions, and that these modules are delivered at the correct levels. Similar to SMART, it should be noted that MOST does not directly assess the overall effectiveness of the intervention (this occurs separately in step 4) to a comparative treatment or control condition. However, the process does ensure that the most efficacious version of the intervention goes forward to this step, thereby making for more efficient use of time and resources.

2.5. Continuous Evaluation of Evolving Intervention Technologies (CEEBIT)

A further recent development in research designs in this field involves the use of CEEBIT (Mohr, Cheung, Schueller, Brown, & Duan, 2013). CEEBIT may be of particular use to clinicians wanting to evaluate multiple versions or ongoing modification of an e-mental health intervention. It allows for multiple versions of web-based or mobile interventions to be deployed simultaneously, with clients randomised as to which version they receive. Versions are then removed if they meet an apriori criteria for inferiority to another version of the intervention. Inferiority is examined rather than superiority as it allows for speedier removal of poorer performing Apps, thus providing added protection for consumers. The strategy allows interventions to be continuously updated and evaluated, promoting ongoing development and assessment of the e-mental health interventions after deployment.

2.5.1. An Applied Example. Assume that following the initial evaluation of PsychAssist (through development, pilot testing or trial) the decision is made to deploy the App to a wider audience, either through open public access or for use

within designated health services. The App is deployed but ongoing revisions to the App continue. These revisions are designed to maintain the functionality of the App with upgrades of operating systems, as well as to provide ongoing improvements and expansion of usability, engagement, and therapeutic content. These revised versions of the App are deployed in an ongoing fashion, with users randomly allocated as to which version of the PsychAssist they receive.

Based on recommendations by Mohr et al. (2013), a 50% alpha rate is employed (rather than the traditional 5%) to determine inferiority of any one version of PsychAssist. This alpha rate is based on the assumption that there will be symmetry of preference among the deployed versions, that is, each is equally likely to be preferred. In contrast, traditional protocols assume preference for the standard research arm and thus require a greater weight of evidence (employing alpha of .05) to demonstrate preference of the new intervention over the standard arm. The 50% alpha level also results in a smaller required sample size for comparing versions of the App, enabling speedier testing, dissemination, and elimination of poorer performing versions of the App.

To protect consumers against prolonged use of inferior Apps, an apriori rule is established to determine when versions will be eliminated. Rather than requiring inferiority of one version of PsychAssist to be demonstrated against all other deployed versions of the App, we decide that if a version demonstrates inferiority (as evidenced by outcomes on the target symptoms of depression or anxiety on the DASS21) against at least one other version of the App, the inferior version will be eliminated. This is deemed to be the most conservative approach to protecting consumers. It should be noted that although we have selected a clinical outcome to determine inferiority, other outcomes such as usage or satisfaction could also be

selected as the outcome measure for assessing inferiority. The selection of the outcome measure may depend on whether the research questions primarily pertain to the efficacy or effectiveness of the App.

2.5.2. Summary. CEEBIT is a design strategy that enables multiple versions of technological interventions to be examined simultaneously (Mohr et al., 2013). The approach tests for inferiority of intervention versions, and allows for timely elimination of poorer performing versions. CEEBIT is particularly congruent with the evolving nature and speed of development of technological interventions. It may also promote the ongoing development and revision of these interventions, fostering continual improvement in the area.

3. Conclusions

It is clear that the rapid development of e-mental health requires use of more flexible and time-effective research designs than offered by the traditional RCT. Use of the strategies presented in this paper may assist clinicians to offer their clients receive mental health interventions that are empirically supported and that are subject to ongoing evaluation and monitoring. Furthermore, although the current paper has focused on the use of these strategies in cognitive behavioural (CBT) interventions for depression and anxiety, it should be noted that the applicability of these designs goes beyond the therapeutic approaches and target populations discussed. The designs discussed in this paper hold considerable promise for answering a range of treatment and delivery questions, regardless of approach (CBT, acceptance and commitment therapy etc.) or condition (depression, anxiety, substance use, etc.). Research has indicated that clients are ready and wanting to engage in these interventions, particularly by means of mHealth technologies. To that end, as scientist-practitioners,

we need to “catch up” with our clients, and provide them with safe and effective interventions.

Funding

This research was funded by a publication assistance scholarship provided by Griffith University, Australia.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

Almirall, D., Nahum-Shani, I., Sherwood, N. E., & Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine, 4*(3), 260-274. doi: 10.1007/s13142-014-0265-0

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs: Uses in applied clinical research. *Archives of General Psychiatry, 29*(3), 319-325. doi: 10.1001/archpsyc.1973.04200030017003

Ben-Zeev, D., Kaiser, S. M., Brenner, C. J., Begale, M., Duffecy, J., & Mohr, D. C. (2013). Development and usability testing of FOCUS: A smartphone system for self-management of schizophrenia. *Psychiatric Rehabilitation Journal, 36*(4), 289-296. doi: 10.1037/prj0000019

Bengali, M. K., & Ottenbacher, K. J. (1998). The effect of autocorrelation on the results of visually analyzing data from single-subject designs. *American Journal of Occupational Therapy, 52*(8), 650-655.

Busse, R., Kratochwill, T. R., & Elliott, S. N. (1996). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33*(4), 269-285.

Casey, L. M., Joy, A., & Clough, B. A. (2013). The Impact of Information on Attitudes Toward E-Mental Health Services. *Cyberpsychology, Behavior, and Social Networking, 16*(8), 593-598. doi: 10.1089/cyber.2012.0515

Casey, L. M., Wright, M.-A., & Clough, B. A. (2014). Comparison of Perceived Barriers and Treatment Preferences Associated with Internet-Based and Face-to-Face Psychological Treatment of Depression. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL), 4*(4), 16-22.

Clough, B. A., & Casey, L. M. (2011a). Technological Adjuncts to Enhance Current Psychotherapy Practices: A Review. *Clinical psychology review, 31*(3), 279-292. doi: 10.1016/j.cpr.2011.03.006

Clough, B. A., & Casey, L. M. (2011c). Technological Adjuncts to Increase Adherence to Therapy: A Review. *Clinical psychology review, 31*, 697-710. doi: 10.1016/j.cpr.2011.03.006

Clough, B. A., & Casey, L. M. (2015a). The Smart Therapist: A Look to the Future of Smartphones and mHealth Technologies in Psychotherapy. *Professional Psychology: Research and Practice*, No Pagination Specified. doi: 10.1037/pro0000011

Clough, B. A., & Casey, L. M. (2015e). Therapy on the move: The development of a therapeutic Smartphone application. *Cyber Behavior, Psychology and Learning, 5*(1), 33-41. doi: 10.4018/ijcbpl.2015010103

Collins, L. M., Baker, T. B., Mermelstein, R. J., Piper, M. E., Jorenby, D. E., Smith, S. S., . . . Fiore, M. C. (2011). The multiphase optimization strategy for engineering effective tobacco use interventions. *Annals of behavioral medicine, 41*(2), 208-226.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American journal of preventive medicine, 32*(5), S112-S118.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research, 15*(11). doi: 10.2196/jmir.2791

Faith, M. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research.

Glasgow, R. E., Magid, D. J., Beck, A., Ritzwoller, D., & Estabrooks, P. A. (2005). Practical clinical trials for translating research to practice: design and measurement recommendations. *Medical Care, 43*(6), 551-557.

Griffiths, K., Farrer, L., & Christensen, H. (2007). Clickety-click: e-mental health train on track. *Australasian Psychiatry, 15*(2), 100-108.

Ioannidis, J. A. (1998). EFfect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA, 279*(4), 281-286. doi: 10.1001/jama.279.4.281

Kazdin, A. E., & Kopel, S. A. (1975). On resolving ambiguities of the multiple-baseline design: Problems and recommendations. *Behavior Therapy, 6*(5), 601-608.

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., . . . Spruijt-Metz, D. (2013). Mobile health technology evaluation: the mHealth evidence workshop. *American journal of preventive medicine, 45*(2), 228-236.

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized medicine, 8*(2), 161-173.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*. Sydney: The Psychology Foundation of Australia.

Manolov, R., Guilera, G., & Sierra, V. (2014). Weighting strategies in the meta-analysis of single-case studies. *Behavior Research Methods, 46*(4), 1152-1166.

Mohr, D. C. (2009). Telemental health: Reflections on how to move the field forward. *Clinical Psychology: Science and Practice, 16*(3), 343-347.

Mohr, D. C., Cheung, K., Schueller, S. M., Brown, C. H., & Duan, N. (2013). Continuous evaluation of evolving behavioral intervention technologies. *American journal of preventive medicine, 45*(4), 517-523.

Nock, M. K., Michel, B. D., & Photos, V. I. (2007). Single-Case Research Designs. In D. McKay (Ed.), *Handbook of Research Methods in Abnormal and Clinical Psychology*. Thousand Oaks, CA: Sage Publications.

Olff, M. (2015). Mobile mental health: a challenging research agenda. *2015, 6.* doi: 10.3402/ejpt.v6.27882

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*(4), 303-322.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behav Ther, 42*(2), 284-299. doi: 10.1016/j.beth.2010.08.006

Riley, W. T., Glasgow, R. E., Etheredge, L., & Abernethy, A. P. (2013). Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. *Clinical and translational medicine, 2*(1), 1-6.

Riley, W. T., Serrano, K. J., Nilsen, W., & Atienza, A. A. (2015). Mobile and wireless technologies in health behavior and the potential for intensively adaptive interventions. *Current Opinion in Psychology, 5*(0), 67-71. doi: http://dx.doi.org/10.1016/j.copsyc.2015.03.024

Rivera, D. E., Pew, M. D., & Collins, L. M. (2007). Using engineering control principles to inform the design of adaptive interventions: A conceptual introduction. *Drug and alcohol dependence, 88,* S31-S40.

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, *52*(2), 109-122.

Shortreed, S. M., Laber, E., Lizotte, D. J., Stroup, T. S., Pineau, J., & Murphy, S. A. (2011). Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Machine learning, 84*(1-2), 109-136.

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling Up mHealth: Where Is the Evidence? *PLoS Medicine, 10*(2), e1001382. doi: 10.1371/journal.pmed.1001382

Zucker, D. R., Ruthazer, R., & Schmid, C. H. (2010). Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *Journal of clinical epidemiology, 63*(12), 1312-1323.